"OPEN-WASHING" IN RUSSIAN OPEN GOVERNMENT DATA

Berezhnaya V.
Associate Professor, "Data Journalism" Programme Academic Supervisor at HSE
(Moscow, Russia)
tinka.ru@gmail.com

Zdorov M.
Student of the Master's Programme "Data Journalism" at HSE
(Moscow, Russia)
3372075@gmail.com

Abstract:
The publication of open government data is aimed at improving quality of life by simplifying access to information, but such is not possible for open-washing data, a kind that is falsely declared as open for everyone. Open-washing data cannot be used effectively due to intentional distortion by the data owners. Christian Villum first drew attention to the worldwide persistence of this phenomenon. Up to today, Russian open government data has not been considered in the context of open-washing, which is critically important now on account of the abolition of the Open Government and other organizations that oversee government openness in Russia. To check if the problem of open-washing is characteristic of Russian open government data, we proposed five quality criteria in this study, intentional noncompliance with which we considered an indicator of open-washing data. Analyzing Russian legislation for compliance with these criteria, we demonstrated it was imbalanced and incomplete. These shortcomings affected the open data sets published by the federal executive bodies, in that most of them can be classified as sets of open-washing data. As a result, we made recommendations based on suggested automation and data stream centralization to improve the quality of the Russian state's open data.
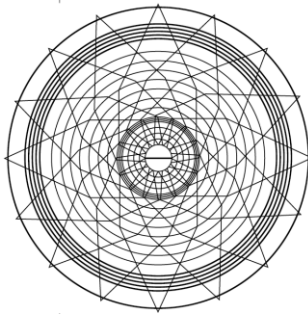**Keywords:** open government data, open-washing, data quality, open government, quality metrics

Introduction

Open data is "the data that governments, legal entities, and individuals provide for wide access, use and exchange" (Open Data Institute, 2015). Its key feature is machine readability, a property by which one can easily access, process, and modify data and its individual elements by a computer (Open Definition, n.d.).

Open data publishing began in the middle of the 2000s and was aimed at increasing the welfare of society by improving the availability of information created by government organizations.

Because the institutional structure of working with governmental open data has already been formed, the main subject of research has been data quality, which is especially important as data has become the basis for management decisions and

business product development. Any undetected flaw (incomplete information, inconsistency with the description, etc.) can lead to incorrect conclusions and considerable losses. Moore (2018) demonstrated in a study that organizations attribute losses of $15 million per year on average to poor quality of available data.

Under these conditions, the term open-washing data, proposed by Danish International Community Manager for the Open Knowledge global network Christian Villum, has been gaining popularity. The term characterizes data whose quality the owners deliberately damaged to render it useless. Villum (2014) noted that the tendency to open-washing is growing, and it is important to recognize the problem and make strategic decisions.

Despite the common use of the term, it has not been considered in the context of Russian open government data, although the question of its quality is more critical than ever: Since the abolition of the Open Government and the Open Data Council in 2018 (Decree of the President of the Russian Federation No 215, 2018), normative and methodological documents have been implemented without support mechanisms and monitoring tools. Thus, it is especially important today to identify existing shortcomings and offer solutions to rework Russian open government data.

In this study, we answered whether the problem of open-washing is typical for Russian open government data and proposed some steps to improve its quality. The article contains specific conditions that can be used in further studies for classifying open government data as open-washing.

Methodology

Because Villum's ideas were framed only in a term, this article proposes conditions for classifying open data as open-washing. Based on analysis and generalization of Russian and foreign academic literature, we determined quality criteria for open data sets where intentional noncompliance might lead to an open-washing problem.

Further, we determined whether Russian legislation that regulates the publication of open data by federal executive bodies creates conditions for aggravating the problem. We considered the legislation's provisions and both the meta-information and contents of Russian federal open data sets.

The exclusion of regional and municipal levels from the research should not have affected its results, as the activities of federal authorities in this field are more regulated and institutionalized enough for the analysis.

This methodology allowed us on one hand to complement the existing ideas of Villum with a practical tool — the condition for classifying the data as open-washing — and on the other to assess the quality of Russian open government data and propose solutions for its improvement.
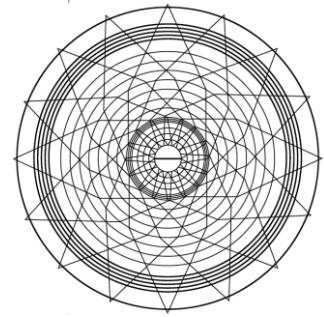
Open-Washing Criteria

The research area devoted to data quality is based on the statement that high-quality data corresponds to its intended use in analytics, decision-making, and planning (Nikiforova, 2018). Throughout its development, the quality of available data

has not improved, although theoretical and practical tools for testing it have been proposed. It is a significant achievement in the existing framework, where data preparation for analysis is the most labor-intensive process: specialists spend 80% of their time searching for data, cleaning and organizing data and only 20% directly on analysis (Gabernet, Limburn, 2017; Pyle, 1999).

The objects of studies in this area have been data and the resources on which it is published. Some methodologies can also be applied to evaluate legislation designed to ensure data quality.

Total Data Quality Management is one of the basic theories focusing on data. It claims that data quality can be evaluated in different directions important to users. Most existing studies suggest these directions (usually from four to ten) and justify their applicability in certain cases (Batini et al., 2009; Bovee et al., 2003; Ferney et al., 2017; Redman, 2001). These directions can be calculated manually or by software such as RapidMiner (n.d.), Microsoft Data Quality Services (2017), and so on.

We assess the resources where open data is published by similar methods. The methods suggest using quality indicators of both the data sets and the development and sustainability of resources (Caro et al., 2007; Neumaier, 2015).

Some software products also automate assessment of resources, the most well-known of them being the Portal Data Quality Assessment Tool (PoDQA tool) and a programme for checking portals built on the CKAN platform (e.g., open data portals of the USA, Canada, and the European Union) (CKAN, n.d.).

Russian theory and practice pay less attention to methodologies for assessing the quality of open government data and the resources on which it is published. The most-used methodology is the state Quality Index of the work with open data of federal executive bodies.

The main limitation of the methodology is the impossibility of an independent assessment of data by users. The index is formed based on questionnaires filled in annually by federal executive bodies. In addition, it does not evaluate the structure or content of data sets – only their availability.

Until 2018, the automated information-monitoring system of the official websites of government bodies as well as local authorities of the Ministry of Economic Development of Russia (AIS "Monitoring of State Sites", n.d.) published daily technical ratings of the work of state bodies with open data. They calculated the rating based on an assessment of the technical characteristics of resources (portals) on which open data was being published. As of mid-2019, they were no longer performing this assessment (AIS "Monitoring of State Sites", n.d.).

The Infometer Project Center annually carries out an independent assessment of open data of federal executive bodies. Unlike the state Index, it assesses the formats and contents of data sets as well as the quality of their metadata (85 indicators in total) (Project Center "Infometer", 2018). A shortcoming of the audit is that it, like the state one, is static and updated only once a year.

Thus, although researchers abroad have gained a great amount of theoretical and practical experience in assessing the quality of data sets and resources where they
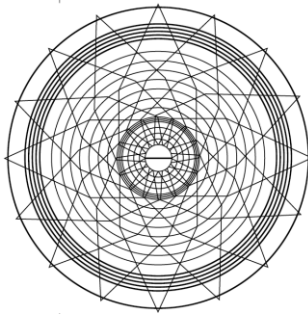
are published, similar methodologies have just begun to develop in Russia. Unlike the foreign ones, these methodologies are based on the goal not of providing users with comprehensive information on data, but of evaluating the work of state bodies.

Because the purpose of this study was to evaluate the quality of the Russian state's open data from the point of view of its users, we mainly used foreign-proposed measurements and indicators. We watered down unified and summarized names and descriptions to five main criteria for evaluating the data quality:

- completeness, or the ability to effectively use the data for its declared purpose;
- timely updates, or the publication of new versions of data as they actually appear;
- reliability, or the compliance of data with its description (content, structure, time range, etc.);
- machine readability, or compliance with generally accepted standards of machine-readable data formats (CSV, RDF, JSON, CSV, XML, etc.);
- openness, or the absence of burdens and limitations such as copyright, patents, and other control mechanisms.

Noncompliance with each of these criteria makes it difficult or impossible to use data. It seems that an open data set can be classified as open-washing if at least one of the criteria is not met, due to the intentional action or inaction of the data owner.

Open-Washing in Russian Open Government Data
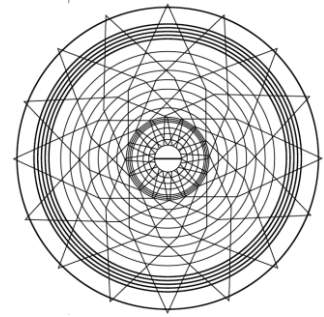***Legislative Requirements***
There are at least seven regulatory legal acts that govern the requirements for Russian open government data (not including internal regulatory legal acts). These acts can conventionally be divided into two groups: those fixing technical requirements (four acts: Decree of the Government of the Russian Federation No 583, 2013; Guidelines for the publication of open data version 3.0, 2014; Order of the Ministry of Communications of Russia No 149, 2013; The Government Commission for the Coordination of Open Government Activities No 7, 2016) and those fixing information that should be published by federal executive bodies (three acts: Decree of the Government of the Russian Federation No 583, 2013; The Government Commission for the Coordination of Open Government Activities No 7, 2016; The Government Commission for the Coordination of Open Government Activities No 10, 2014).

By analysis of the technical requirements, we showed that they unify the external presentation of data sets and the structure of their meta-information, consolidate the general requirements for openness, and declare the principles of machine readability, completeness, reliability, and relevance. However, the requirements do not contain tools instrumental to the implementation of these principles, such as formats for automatic downloads from departmental information systems, encodings, excerpts from standards of machine-readable formats, update periods, and so forth.

The regulatory legal acts fixing the information that should be published by federal bodies contain 407 open data sets. To assess the quality of these requirements, we developed a methodology covering two levels:
- completeness of the requirement, and
- the demand for the required open data set.

On the first level, we assessed how fully each requirement is formulated and whether it stimulates the qualitative publication of an open data set. We checked the following criteria:
- The structure of the open data set is fixed. Without this condition, similar sets published by different state bodies would have incompatible structures. In addition, information in open data sets may be overly generalized or irrelevant. If "Yes," one point was given; otherwise, zero points were given.
- A period for updating the open data set is specified. If "Yes," one point was given; otherwise, zero points were given.
- The requirement for publishing an open data set is unique. Duplication of requirements complicates legislation. It is more difficult for state bodies to figure out what they should publish and how to report on it correctly. If "Yes," one point was given; otherwise, zero points were given.

At the first level, a requirement could score up to three points. Only when all three criteria are met do we consider the wording of the requirement to be complete.

On the second level, we assessed if an open data set is in demand for analysts and programmers. We checked the following criteria:
- The data often changes. This does not imply a regulatory deadline for updating an open data set, which we mentioned earlier, but a description of the information itself. Data that changes less often than once per month will not be needed in a machine-readable format, as it is too paced out to develop a service, application, or conduct research on its basis. Such data may be published in other sections of the official website of the state body. If the data is updated at least one time per week, then one point was given; if it is updated at least one time per month, then 0.5 points were given; if it is updated less than one time per month, then zero points were given.
- The data is big. Similarly to paragraph above, if the dataset contains more than 1,000 rows, one point was given; if it contains fewer than 1,000 but more than 100 rows, then 0.5 points were given; and if it contains fewer than 100 rows, then zero points were given.
- The data is quantitative, which is required for presentation in tabular form. If "Yes," one point was given; otherwise, zero points were given.

At the second level, a requirement could score up to three points. Upon reaching 2.5 points, we considered a set of open data as needing analysts and programmers.

Thus, we considered a requirement to be qualitative if it scored 5.5 points across both levels. We identified only 15 out of 407 (3.7%) open data sets as qualitative.
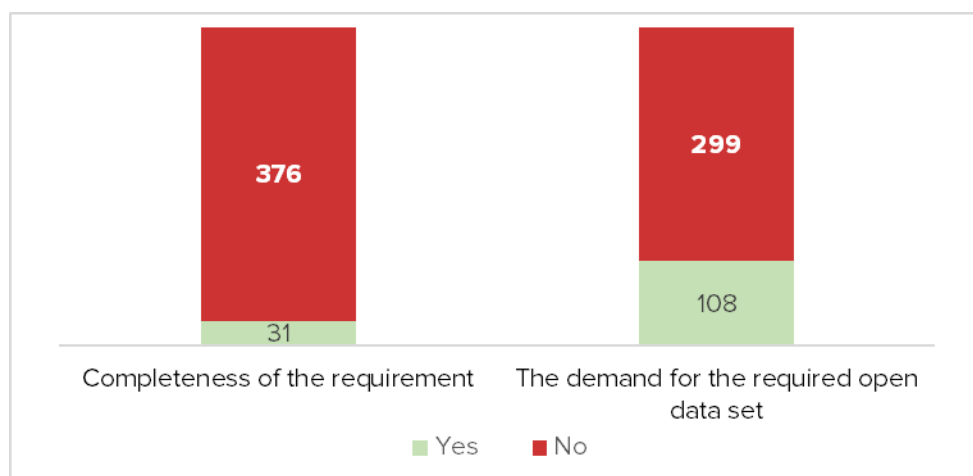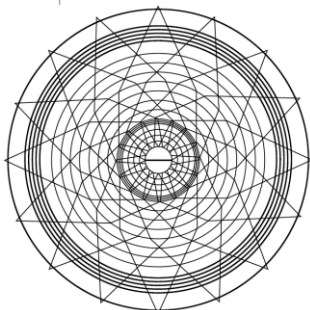
Figure 1. Compliance of the sets with the criteria of two levels of analysis, pcs

Figure 2. Sets that meet the criteria of two levels, pcs

At the first level, we identified 31 out of 407 (7.6%) requirements as complete. The main problem is that 359 requirements (88%) of the set structure are not formalized.

Of the requirements, 260 (64%) do not fix a deadline for updating them. Only one out of every three regulatory legal acts contain an updated period. At the same time, each of these acts requires updating a significant part of the sets "when changes appear," so the deadline is not specific.

We also revealed that 25 requirements (6%) are not unique. For example, the following requirements actually duplicate each other:

- "names of subordinate organizations (if any)";
- "a list of subordinate organizations (if any), information about their tasks and functions, as well as postal addresses, email addresses (if any), phone numbers of the help services of subordinate organizations";
- "a list of educational institutions subordinate to the federal executive body (if any), indicating their mailing addresses, addresses of official websites, as well as phone numbers by which reference information about these educational institutions can be obtained".
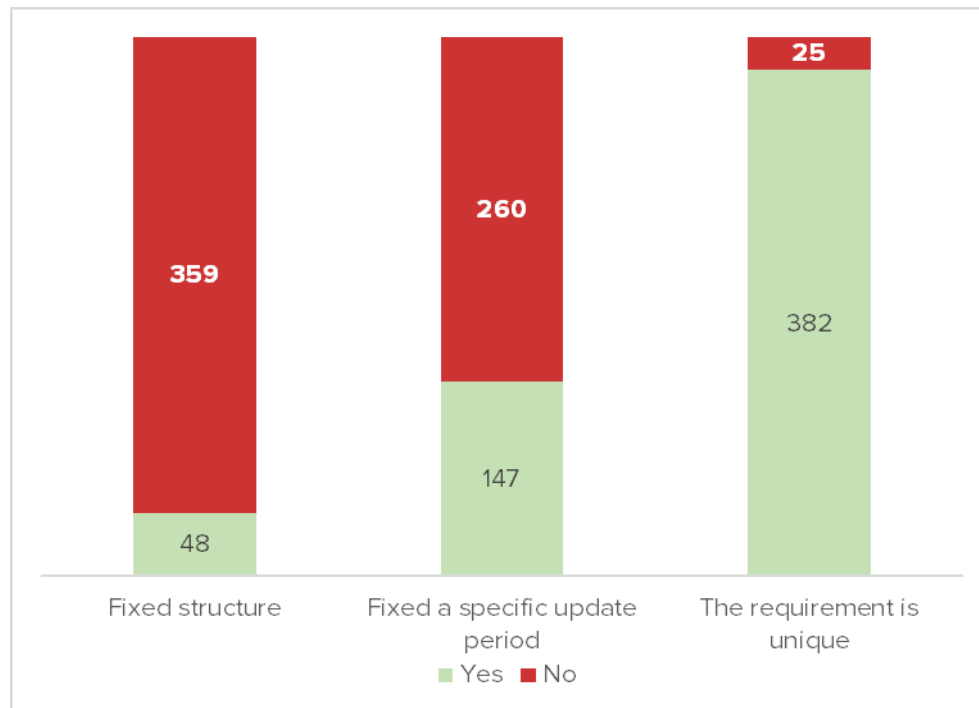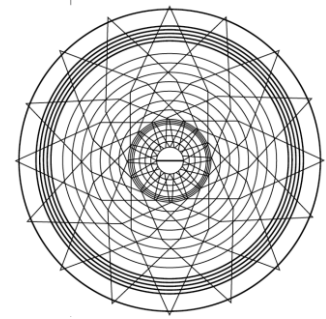
Figure 3. Compliance of the sets with the criteria of completeness, pcs

In the second-level assessment, we identified only 108 out of 407 sets (26.5%) as demanded by data consumers. The main problem is that only 117 sets are regularly updated (28.7%). For example, the data for the sets "Plan of scheduled inspections of legal entities and individual entrepreneurs for the next year" and "List of public services" is changed no more than once per year. In the first case, the plan is approved annually, and inspections outside of it are unscheduled. As for the list of public services, changing it requires the approval of Russia's Ministry of Economic Development and Ministry of Justice, which takes at least a year to procure (Federal Law No 210-FZ "On the Organization of the Provision of State and Municipal Services", July 27, 2010).

We also found data sets containing a small amount of data — up to 1,000 (155, 38%) or even 100 rows (142, 35%). Examples of such sets include "The list of state information systems administered by the federal executive body, its territorial bodies and its subordinate organizations," "Russia's position in the ranking of ease of doing business, according to the World Bank Doing Business," and so forth.

In addition, we identified 21 sets (5%) that had contents not suitable for tabular display. Examples are "The procedure for appealing regulatory legal acts and other decisions adopted by a state body, its territorial bodies, municipal legal acts," "Information on the participation of the federal executive body in international cooperation, including the official texts of international treaties of the Russian Federation," and so forth.
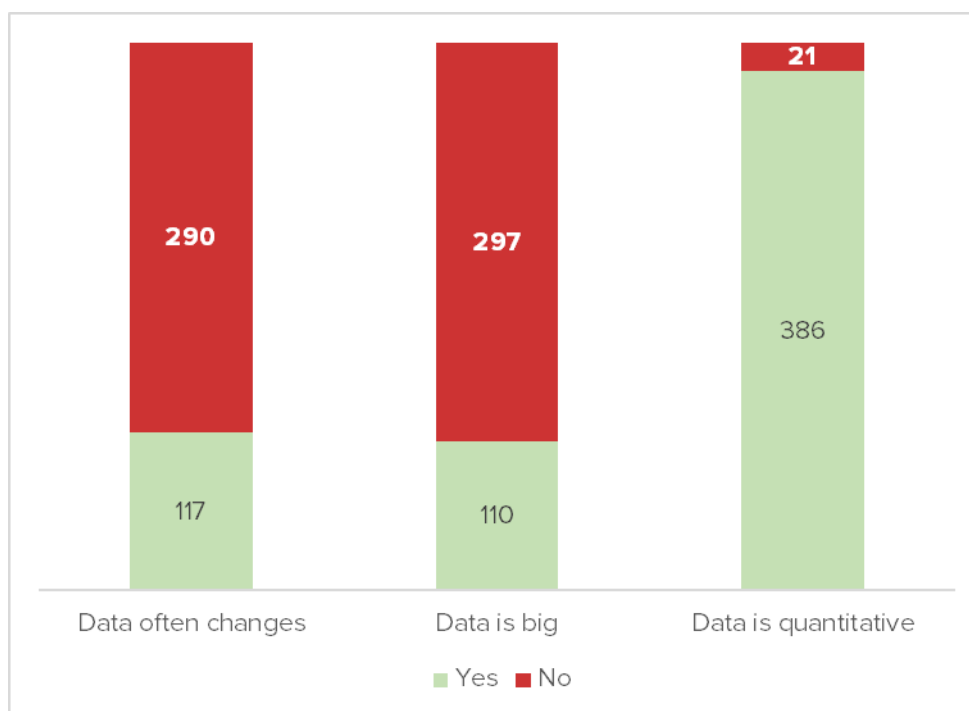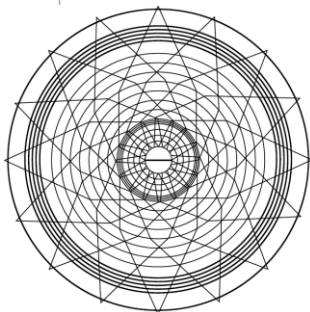
Figure 4. Compliance of the sets with the criteria of demand, pcs

At the same time, these names are stated in the Decree of the Government of the Russian Federation No. 953 (2009) and the federal law "On ensuring access to information on the activities of state bodies and local governments" No. 8-FZ (2009) for posting on the websites of federal bodies in a human-readable form. However, the methodology for assessing organizational and technical readiness and the quality of the organization of work with open data (Open Government Knowledge Base, n.d.; The Government Commission for the Coordination of Open Government Activities No 7, 2016) obliged federal bodies to publish information as open data sets in tabular form for which the data are not suited.

Analyzing the legislative requirements for Russian open state data, we showed their imbalance. On one hand, the list of open data sets that should be published by federal executive bodies has been excessively regulated. It seems that most of the sets will be uninteresting for analysts and programmers. At the same time, federal bodies cannot delete an unclaimed data set, as it must be posted.

On the other hand, little attention has been paid to data quality. Machine readability, completeness, timely updates, and reliability are only declared, but there are no mechanisms for their implementation or monitoring. Simultaneously, detailed technical requirements have been approved to ensure that the openness criterion is met.

Because four out of five criteria are not sufficiently developed in Russian legislation, an open data set can be deliberately and legally published with errors that prevent its effective use. Thus, there are no tools for protection against open-washing in Russian legislation.
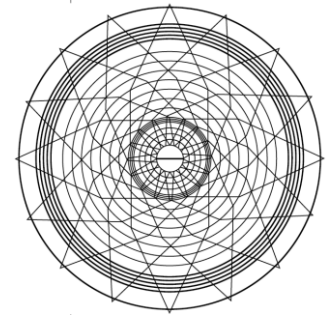
Under these circumstances, we suggest that criteria for machine readability, completeness, timely updates, and reliability will not be met for a significant part of the Russian open government data sets. The following hypotheses can be formulated for these sets:

- They are unclaimed;
- They are irrelevant;
- They have shortcomings associated with structure, including incomplete recruitment structures that do not disclose their declared content and structures of the same names of open data sets placed by federal executive bodies unique and incompatible with each other;
- They have drawbacks associated with format, including export errors, encoding errors, and errors in using standards of machine-readable formats.

The quality of published Russian open government data

We identified 75 resources on which federal executive bodies place their open data sets, consisting of

- 56 open data sections on their official websites (for the Ministry of Natural Resources of Russia, the Ministry of Health of Russia, the Ministry of Transport of Russia, etc.),
- 7 open data portals of federal executive bodies (for the Ministry of Culture of Russia, Rostourism, Rosmorrechflot, etc.), and
- 12 other open data portals for which the federal executive bodies are responsible (the Open Data Portal of the Russian Federation, the open data of the Federal State Information System for the Coordination of Informatization, the Unified Procurement Information System, etc.).

The Russian federal executive bodies had published 17,058 open data sets. To test our hypotheses and evaluate the quality of the sets, we developed a methodology covering two levels:

- quantitative assessment based on meta-information (passport and other information on the resources): relevance rate, list of the most popular sets, and the amount of their downloads (if there is a counter on the resource);
- qualitative assessment based on the content of the sets: identification of incomplete and incompatible structures and other shortcomings.

Relevance

At the start of the analysis, we found it would not be possible to determine the relevance of the data sets either of the Federal State Statistics Service (Rosstat) or placed in the Unified Interdepartmental Information and Statistical System (EMISS; 1,746 and 6,181 sets, respectively). Due to the statistical nature of the data, the sets were published once without subsequent updates. When new data appears, a new set is created — for example, "Balance Currency for 2016" before "Balance Currency for 2017."
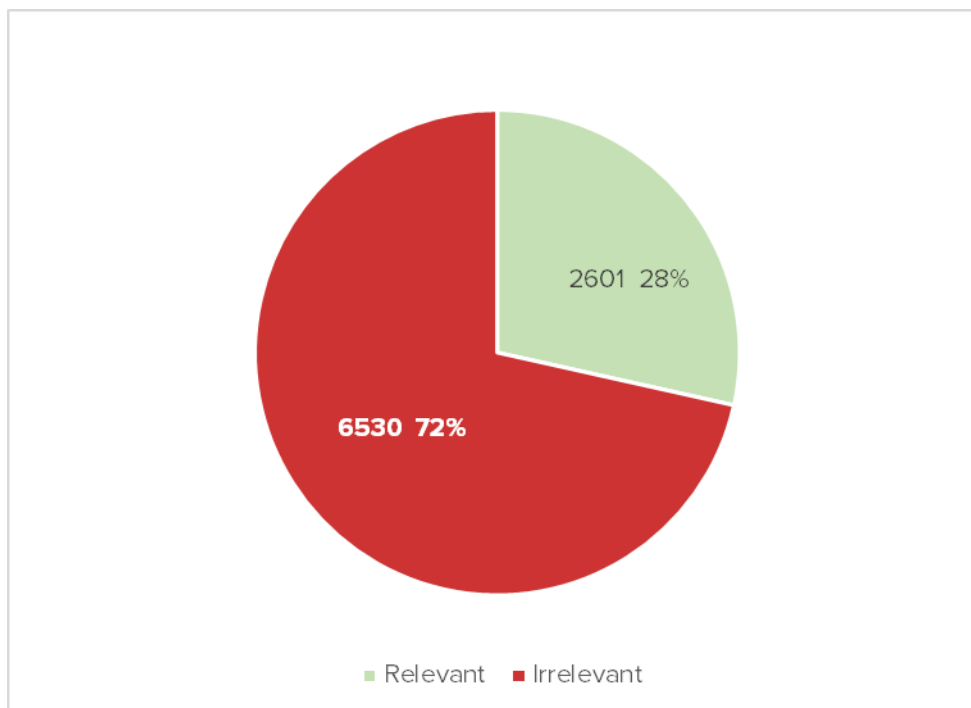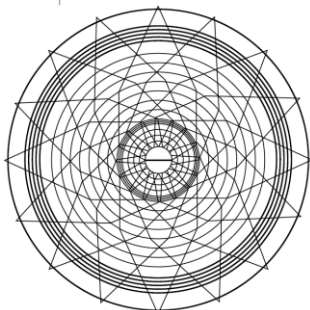
Figure 5. Relevance of federal open data sets, pcs and %

It would be wrong to call a set with statistics of past periods irrelevant because the date of relevance in its passport has passed. This information can also be used. However, in the current analysis, we could not call such sets relevant and thus did not evaluate them. The inconvenience of such a scheme for publishing data should be noted: a large number of data sets appear, making navigation among them difficult and obstructing access to information.

Under these conditions, we identified 2,601 open data sets as relevant, or 28.5% of the analyzed sets (9,131 pieces, minus data from Rosstat and EMISS).

Demand

Of the examined resources, 26 have functionality for evaluating the download statistics of open data sets. On each of them, we analyzed the three most-demanded open data sets. As a result, 29 of the 78 open data sets (37%) were published in compliance with the requirements, and the remaining 49 (63%) were published at the initiative of a federal executive body. Moreover, each compulsory set was downloaded on average (arithmetic mean) 98,315 times, and a set published at the initiative of the body was downloaded 110,720 times.

We found that data from registers, databases, and other information systems of ministries and state agencies are in great demand among users: of the 10 most popular sets, at least 9 can be attributed to this category. Examples include the "Register of notifications on the beginning of certain types of entrepreneurial activities" and "Information from the Unified State Register of Cultural Heritage Sites (historical and cultural monuments) of the peoples of the Russian Federation." Some of these sets are published without human intervention on these information systems.
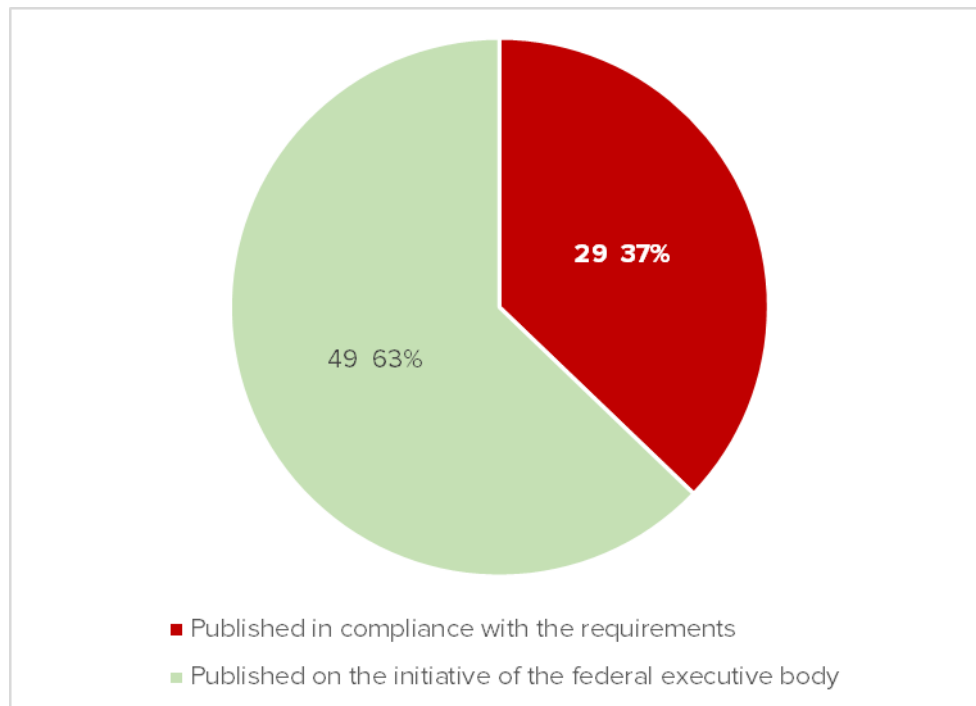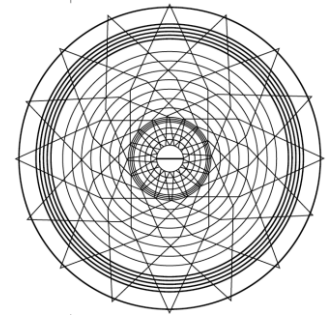
Figure 6. Most-demanded federal open data sets, pcs and %

Completeness of Set Structures

We discovered some episodes of incomplete structures among open data sets of federal executive bodies. For example, based on the name, the set belonging to the Ministry of Finance of Russia and called "Federal property owned by a state authority, including property assigned to subordinate federal state unitary enterprises and federal state institutions" suggests a list of such property. In contrast, the structure represents only these quantities:

- – "the number of real estate assets"
- – "the number of objects of movable assets"

The set "Organizational structure of the Federal Treasury" is also overly generalized. Its structure consists of three positions:

- – "structure of the Federal Treasury"
- – "civil Status"
- – "budget and legal status"

Machine-Readable Set Formats

Similarly, we identified shortcomings in publication format of the sets. The "State Register of Medicinal Products" of the Ministry of Health of Russia, the "Schedule for Public Procurement" of Rosmorrechflot, and "Statistical Information on the Main Activities of the Federal Penitentiary Service of Russia" are published in CSV format with a comma delimiter. However, they do not comply with the CSV RFC 4180 standard in that the data does not contain separate commas and empty-cell values are skipped without using a separator. Under these conditions, the use of sets without pretreatment is impossible.
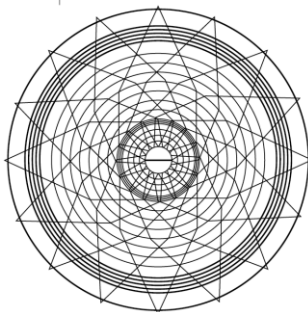
The set "Information on incomes, expenses, on property and property obligations of state civil servants" of Rostrud contains information on both state civil servants and their family members, but the latter individuals are not linked to a specific civil servant. Thus, this set also cannot be processed without human intervention.

The problem of data blinding is also relevant to "Schedule of scheduled inspections" of Rosavtodor. Moreover, in the set, an abbreviation (legend) following the data does not fit into the set's existing structure and may be mistakenly perceived as a continuation of it.

Structural Compatibility

The federal executive bodies must publish open data sets containing lists of their subordinate organizations. The acts fixing this requirement do not contain a description of set structure (Order of the Government of the Russian Federation No. 1187-r, 2013; The Government Commission for the Coordination of Open Government Activities No 7, 2016). Analyzing the legislative requirements for Russian open government data, we identified the data as not in demand for analysts and programmers due to its rare updatability and small size. However, we found that combined data on subordinate organizations of all state bodies would be in demand.

During the audit of the resources, we identified 52 open data sets with information on subordinate organizations. The structures of nine of these (17%) are the same, and the remaining 43 sets contain from one to 22 columns. Thus, most of the similar sets cannot be combined with each other because they have different structures.
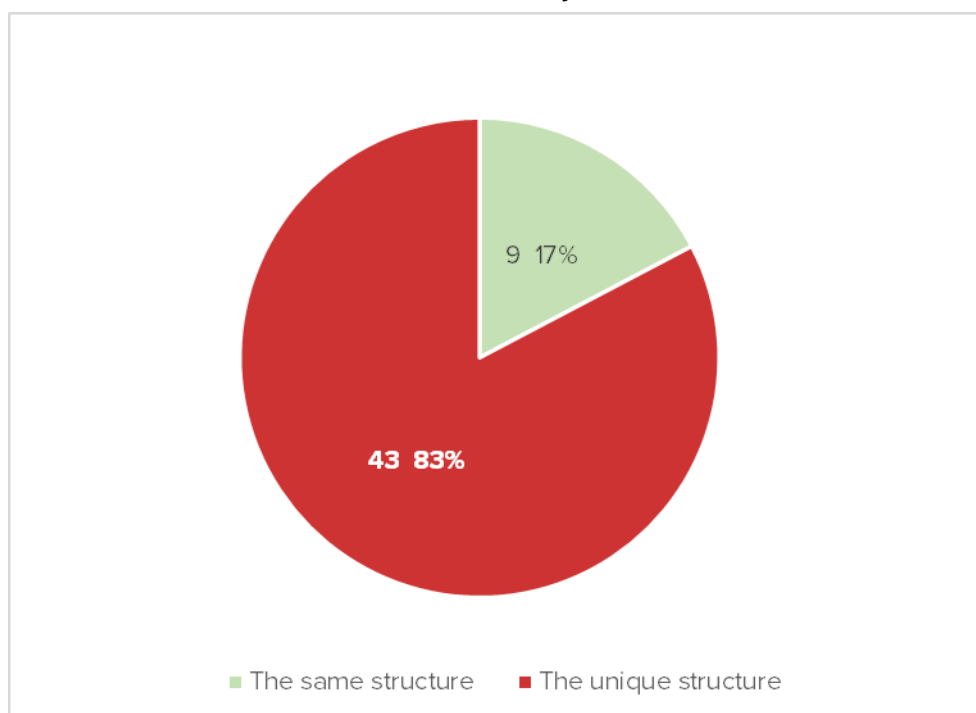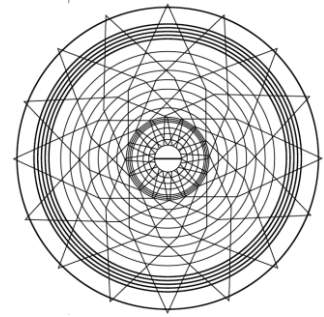

Figure 7. Structural compatibility of the open data sets containing a list of subordinate organizations of the federal executive bodies, pcs and %

Recommendations for Improving the Work of Russian Federal Executive Bodies with Open Data

First, it was necessary to optimize the current list of open data sets required for publication. This could be achieved by eliminating duplication of data centralized in common information resources. For example, we transferred data on inspections planned and carried out by all state bodies, as well as on state services rendered by them, to the state-run automated information system "Management." The Ministry of Economic Development of Russia maintains a list of all public services (Federal Law No. 210-FZ, 2010), and the Federal Tax Service of Russia aggregates information about existing licenses issued by both federal and regional authorities (Federal Law No. 99-FZ, 2011). It seems that one centralized data set would be more valuable for users than many sets of different state bodies that are incompatible with each other due to different formats, structures, and update periods. Thus, the number of data sources could be optimized, and their quality would only improve.

The remaining required sets should be revised due to their uselessness for analysts and programmers. Data owners should also be included in this assessment, as they know the current formats and structures of data accounting.

For new data sets, it is necessary to fix a specific update period, data structure, and automated publication format (publication directly from information systems without the participation of an employee of a state body). We showed that such data is in high demand among users.

Automated publication would require additional investments in the development of IT infrastructure but would ensure the highest quality of data. It would have a stable structure and update period and would comply with generally accepted standards of machine-readable formats.

Secondly, new open data-monitoring mechanisms to assess the quality of the data from the point of view of their users should be introduced. Such assessment should be carried out regularly and preferably in an automated form. Thus, users at any time would be able to verify the quality of the data they use.

Conclusion

As open government data is increasingly being used to make managerial decisions and create business products, attention to its quality is growing. Villum said that owners of data often spoil it to make it harder to use. He called such data manipulation open-washing.

Open-washing often happens when people are required to publish data against their will. Intent to sell qualitative data or fear that in-depth analysis of the data may raise questions to its owners could also be reasons for open-washing.

To determine if this problem is typical for Russian open government data, we defined the conditions for classifying a data set as open-washing is noncompliance with five quality criteria:
- completeness, or the ability to effectively use the data for its declared purpose;
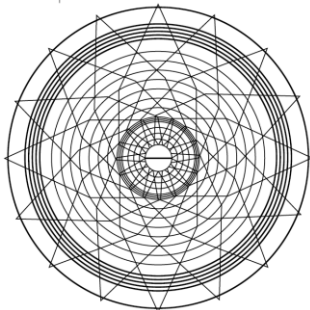
- timely updates, or the publication of new versions of data as they actually appear;
- reliability, or the compliance of data with its description (content, structure, time range, etc.);
- machine readability, or compliance with generally accepted standards of machine-readable data formats (CSV, RDF, JSON, CSV, XML, etc.);
- openness, or the absence of burdens and limitations such as copyright, patents, and other control mechanisms.

As a result of our analysis of Russian open data legislation, we revealed its imbalance. While the list of open data sets that must be published by federal executive bodies is overregulated, almost no attention is paid to the sets' quality. Machine readability, completeness, timely updates, and reliability are merely declared, whereas there are no mechanisms or tools for implementing and monitoring them. At the same time, detailed requirements are given to unify the technical side of the publication, ensuring that the openness criterion is met.

Four of the five proposed quality criteria are not sufficiently developed in Russian legislation, which means that legal conditions are created for the publication of open-washing data. Analyzing the published open data sets, we confirmed these risks:

- out of 9,131 open data sets posted by federal executive bodies on 73 resources, only 29.5% were relevant;
- individual episodes of low-quality data were identified and associated with an overly generalized and irrelevant structure of the sets, as well as with noncompliance with the standards of machine-readable formats;
- we confirmed that similar sets placed by different bodies are incompatible with each other and cannot be used as efficiently as possible.

Thus, the vast majority of federal open data sets cannot be used effectively. Moreover, the presence of incomplete structures and systematic lack of set updates indicate that these mistakes in open data sets were made intentionally. These findings confirm that a significant part of Russian open government data is characterized by the open-washing problem.

This means that Russian open government data has low potential for making effective management decisions, creating new business solutions, and increasing the added value of existing ones. Publishing these data sets, which takes both labor and financial resources, poorly contributes to the development of the Russian welfare. Thus, the data sets do not fulfil the main mission of disclosing government data.

We have developed the following recommendations to improve the work of Russian federal state bodies with open data, which would allow for balancing legislation:

- optimization of data sources: data centralized in a common resource during its life cycle should be published from that resource. This would reduce the burden on state bodies, and one complete set would be published instead of dozens of scattered and incompatible ones;
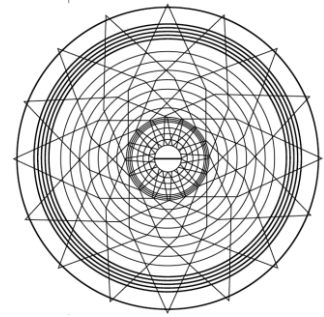
    — review of compulsory publication sets: during this, both data users and data owners should be interviewed.

    — predominant automatic publication of data sets from state information resources: this would almost eliminate manual manipulation of data quality and therefore solve the open-washing problem. Automatically published sets would have a stable structure and update period and would also comply with generally accepted standards of machine-readable formats;

    — open and automatic quality control tools for Russian open government data: these could be implemented and taken advantage of by the data's users.

We believe that this study can be continued to identify the causes of the open-washing phenomenon and the motivations of data owners for the intentional distortion of data.

REFERENCES

AIS "Monitoring of State Sites" (n.d.). https://gosmonitor.ru

Batini, C., Cappiello, C., Francalanci, C., Maurino, A. (2009). Methodologies for Data Quality Assessment and Improvement. ACM Computing Surveys, 41(3), 1–52. https://doi.org/10.1145/1541880.1541883

Bovee, M., Srivastava, R. P., Mak, B. (2003). A conceptual framework and belief‑function approach to assessing overall information quality. International Journal of Intelligent Systems, 18(1), 51‑74. https://doi.org/10.1002/int.10074

Caro, A., Calero, C., Piattini, M. (2007). A Portal Data Quality Model for Users and Developers. 12th International Conference on Information Quality (pp. 462–476). MIT.

CKAN (n.d.). About CKAN. https://ckan.org/about/

DAMA UK Working Group on "Data Quality Dimensions". (2013). The six primary dimensions for data quality assessment. https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf

Decree of the Government of the Russian Federation No 583. (2013, July 10). On Providing Access to Publicly Available Information on the Activities of State Bodies and Local Governments in the Information and Telecommunication Network "Internet" in the Form of Open Data. https://www.rosminzdrav.ru/documents/8169-postanovlenie-pravitelstva-rf-ot-10-iyulya-2013-g-583-ob-obespechenii-dostupa-k-obschedostupnoy-informatsii-o-deyatelnosti-gosudarstvennyh-organov-i-organov-mestnogo-samoupravleniya-v-informatsionno-telekommunikatsionnoy-seti-internet-v-forme-otkrytyh-dannyh
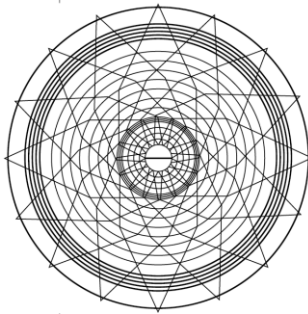
Decree of the Government of the Russian Federation No 953. (2009, November 24). On Providing Access to Information on the Activities of the Government of the Russian Federation and Federal Executive Bodies. http://base.garant.ru/196682/

Decree of the President of the Russian Federation No 215. (2018, May 15). On the Structure of Federal Executive Bodies. https://base.garant.ru/71942842/

Federal Law No 210-FZ. (2010, July 27). On the Organization of the Provision of State and Municipal Services. http://base.garant.ru/12177515/

Federal Law No 8-FZ. (2009, February 9). On Providing Access to Information on the Activities of State Bodies and Local Authorities. http://base.garant.ru/194874/

Federal Law No 99-FZ. (2011, May 4). On Licensing of Certain Types of Activities. http://base.garant.ru/12185475/

Ferney, M., Estefan, L., Alexander, V. (2017). Assessing data quality in open data: A case study. Congreso Internacional de Innovacion y Tendencias en Ingenieria (pp. 1–5). IEEE. https://doi.org/10.1109/CONIITI.2017.8273343

Gabernet, A., Limburn, J. (2017). Breaking the 80/20 rule: How data catalogs transform data scientists' productivity. IBM. https://www.ibm.com/blogs/bluemix/2017/08/ibm-data-catalog-data-scientists-productivity/

Guidelines for the publication of open data version 3.0 (2014, May 29). RF. https://data.gov.ru/metodicheskie-rekomendacii-po-publikacii-otkrytyh-dannyh-versiya-30

Microsoft. (2017). Overview of Data Quality Services. https://docs.microsoft.com/en-us/sql/data-quality-services/introduction-to-data-quality-services?view=sql-server-2017

Moore, S. (2018). How to Create a Business Case for Data Quality Improvement. Gartner. https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/

Neumaier, S. (2015). Open Data Quality: Assessment and Evolution of (Meta-) Data Quality in the Open Data Landscape. Technische Universität Wien. https://sebneumaier.files.wordpress.com/2016/11/master_thesis_presentation.pdf

Nikiforova, A. (2018). Open Data Quality Evaluation: A Comparative Analysis of Open Data in Latvia. Baltic J. Modern Computing, 6(4), 363–386. https://doi.org/10.22364/bjmc.2018.6.4.04

Open Data Institute. (2015). Open data in government: how to bring about change. https://theodi.org/article/open-data-in-government-how-to-bring-about-change/

Open definition. (n.d.). Open Definition 2.1. http://opendefinition.org/od/2.1/en/
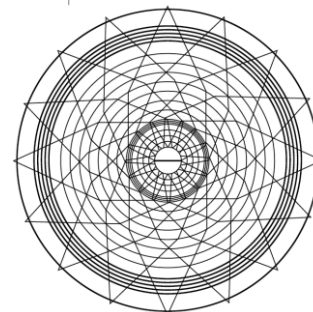
Open Government Knowledge Base. (n.d.). Comparative analysis of legislative requirements for information disclosure. https://clck.ru/NG8C4

Order of the Government of the Russian Federation No 1187-r. (2013, July 10). On the Lists of Information on the Activities of State Bodies, Local Authorities Posted on the Internet in the Form of Open Data.
http://www.garant.ru/products/ipo/prime/doc/70313602/

Order of the Ministry of Communications of Russia No 149. (2013, June 27). On Approval of the Requirements for Technological, Software and Linguistic Means Necessary for Posting Information by State Bodies and Local Authorities on the Internet in the Form of Open Data, as Well as to Ensure Its Use.
http://base.garant.ru/70438250/

Project Center "Infometer". (2018). Data for nobody.
https://read.infometer.org/foivod2018

Public.Resource.Org. (2007). Open Government Working Group Meeting in Sebastopol. https://public.resource.org/8_principles.html

Pyle, D. (1999). Data preparation for data mining. Morgan Kaufmann Publishers.

RapidMiner. (n.d.). About Us. https://rapidminer.com/us/

Redman, T. (2001). Data quality: the field guide. Digital press.

The Government Commission for the Coordination of Open Government Activities No 7. (2016, October 28). The Methodology for Assessing Organizational and Technical Readiness and the Quality of the Organization of Work With Open Data. (quality Index of the Work of Federal Executive Authorities With Open Data).
https://clck.ru/NGdFg

The Government Commission for the Coordination of Open Government Activities No 10. (2014, December 25). The Schedule for Disclosing Priority Socially Significant Data Sets (an Appendix to the Action Plan "Open Data of the Russian Federation").
https://clck.ru/NGdaw

The Open Data Institute. (2017). Recommendations for Open Data Portals: from setup to sustainability. European Data Portal.
https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf

Villum, C. (2014). "Open-washing" – The difference between opening your data and simply making them available. Open Knowledge Foundation.
https://blog.okfn.org/2014/03/10/open-washing-the-difference-between-opening-your-data-and-simply-making-them-available/
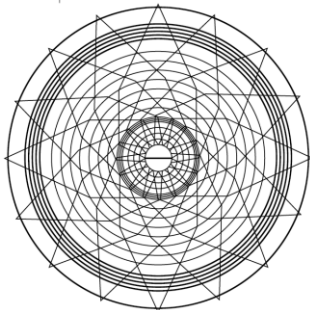
ПРОБЛЕМА «OPEN-WASHING» В РОССИЙСКИХ ОТКРЫТЫХ ГОСУДАРСТВЕННЫХ ДАННЫХ

Бережная В.С.

Доцент, академический руководитель образовательной программы «Журналистика данных» Национального исследовательского университета «Высшая школа экономики» (Москва, Россия)
tinka.ru@gmail.com

Здоров М.Ю.

Студент магистерской программы «Журналистика данных» Национального исследовательского университета «Высшая школа экономики» (Москва, Россия)
3372075@gmail.com

Аннотация:

Публикация открытых государственных данных направлена на улучшение качества жизни и упрощение доступа к информации. Однако данные часто оказываются намеренно испорчены владельцами, чтобы сделать их непригодными для анализа. Это явление получило название «open-washing», впервые предложенное специалистом в области открытых знаний Кристианом Виллумом. Проблема open-washing данных актуальна во всем мире.

До этого момента открытые государственные данные Российской Федерации не рассматривались в контексте open-washing. Сегодня это становится критически важным в связи с роспуском правительственной комиссии по координации деятельности Открытого правительства и других организаций, осуществляющих контроль за открытостью высшего исполнительного органа государственной власти. Чтобы определить наличие проблемы open-washing в российских открытых государственных данных, предложены пять критериев. Намеренное несоответствие этим критериям считается индикатором open-washing данных.

Анализ российского законодательства по предложенным критериям также выявил его неполноту и предвзятость. Недостатки законодательства в свою очередь повлияли на качество данных, опубликованных федеральными исполнительными органами: большинство представленных наборов данных можно классифицировать как open-washing. В конце статьи авторы дают рекомендации по улучшению качества открытых государственных данных, основанные на предложенных системах автоматизации и централизации потоков данных.

Ключевые слова: открытые данные, качество данных, open-washing, проверка качества, метрики качества